DCU ADAPT at TRECVID 2021

Video Summarization - Keeping It Simple

A. Potyagalova, Gareth J. F. Jones

Overview

- Introduction
- Approach
- Main task
- Subtask
- Results and Conclusions
- Questions

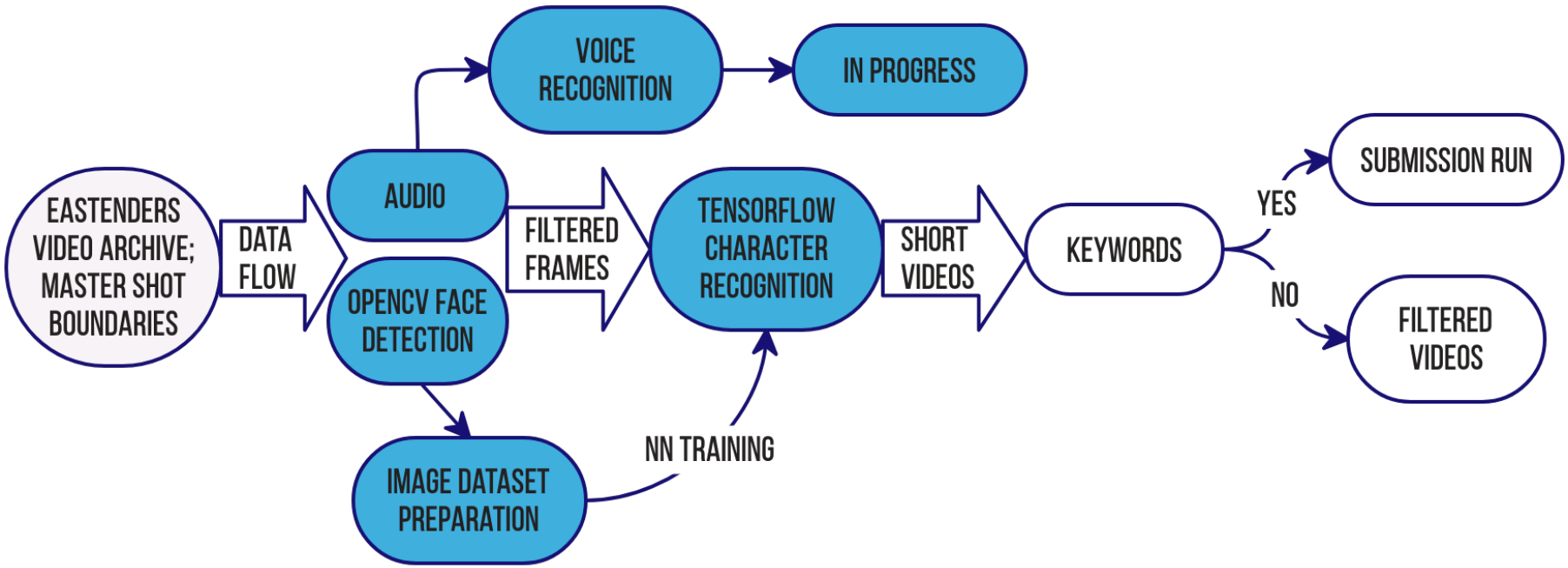
Introduction

Introduction

- **Approach** Main approach to VSUM task detect sub-clips containing selected characters using a neural network:
  - face-detection and keyword search in short clips.
- **Results** Achieves reasonably good result (30.5% accuracy) for main task, but rather poor result for subtask questions (17.2% accuracy).
- **Ways to improve** Handle questions related to individual characters separately; perform more detailed analysis of subscripts.
  Current voice recognition results were not accurate enough to include them in the final submission.

Approach

Approach
Scheme



Approach
Corpus creation

- **Image dataset**

Figure 1: Extraction of frames and audio, and preparation for training datasets

- **Preparation** includes standard Keras augmentation for images; and adding minor white noise to audio chunks

Approach
Corpus creation

- **Metadata extraction** Scraping synopses from video metadata and fansites.
  - Idea is that if a character is not mentioned in the episode synopsis, there will be no important events for that character.
- **Keyword storage** Creation of keyword list for detection of major events
  - Idea is that specific keywords can serve as a flag to determine the importance of the episode.

Main task

Main task

Neural network training

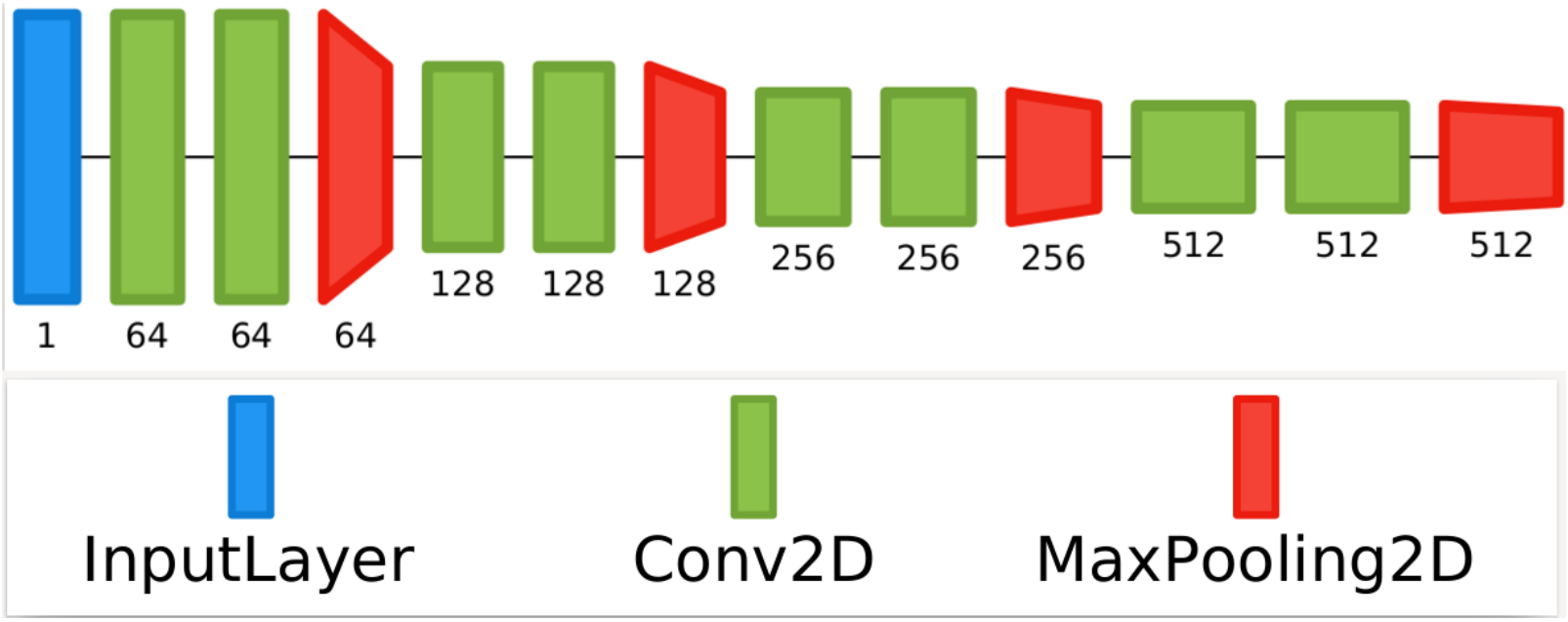


Figure 2: Tensorflow-based CNN. Regularization methods from the Keras API were added to solve overfitting problem.

Main task

Detect character

- **Character recognition**

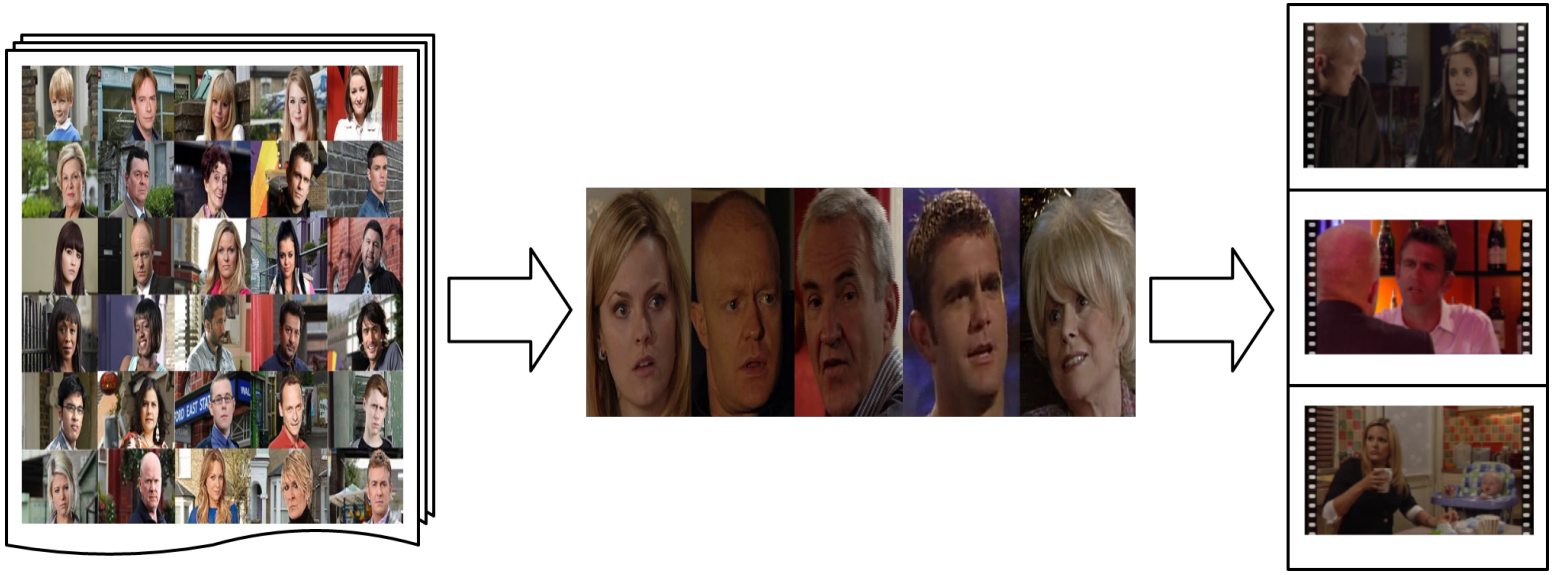


Figure 3: Detect all faces in video, select 5 listed characters, and extract relevant subclips

- **Subclip Set** Creates a set of subclips containing the selected characters.

Main task

Video processing

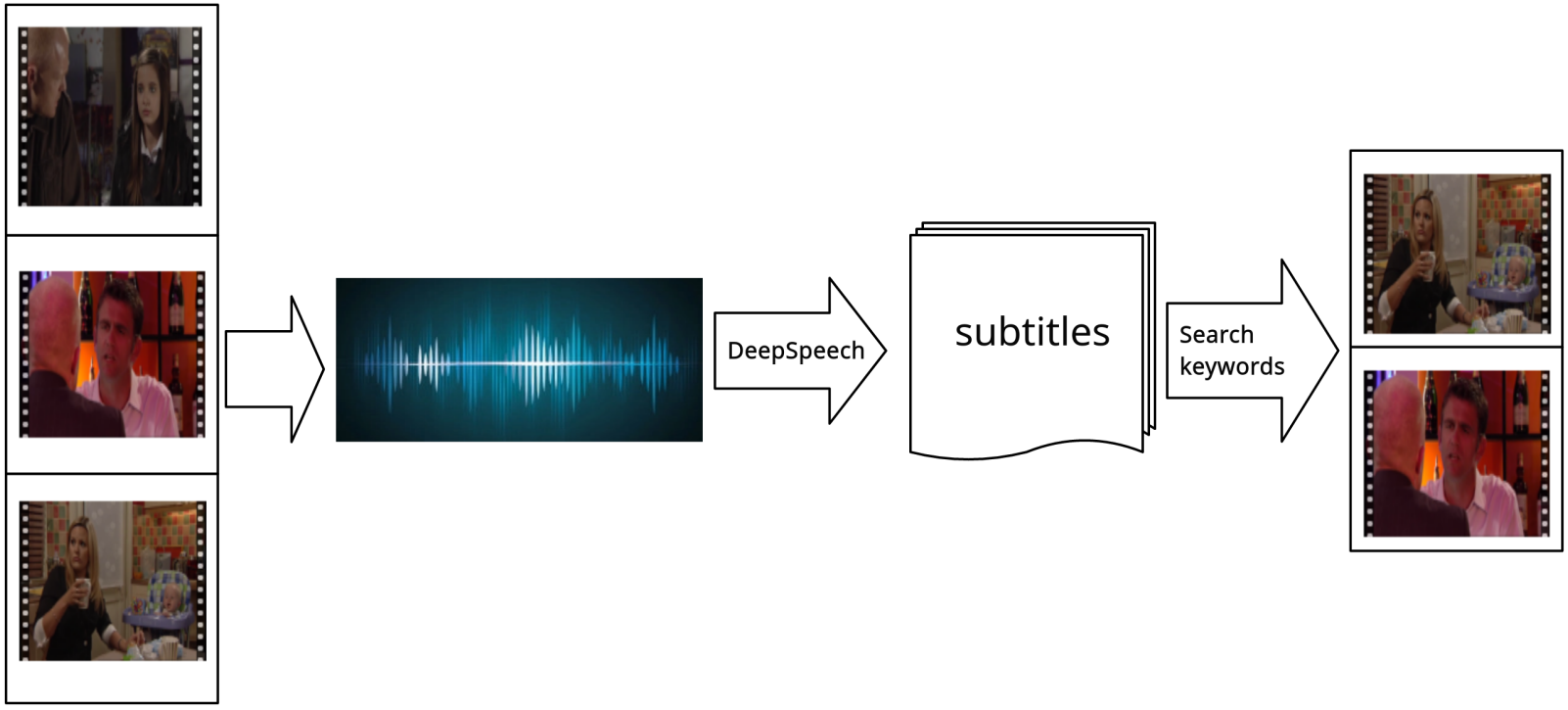


Figure 4: Audio track extracted from the clip, speech transcribed using Deep-Speech. Keywords searched in the text file of each clip.

Subtask

Subtask

Subtask

- **Keyword lists** Separate pool of keywords created for each question.
- **Search** The same search stages used as in a main task.

Results and Conclusions

Results and Conclusions

Results and scores

Main task results	
Query	Percentage
Adapt_Archie	50.50%
Adapt_Jack	19.25%
Adapt_Max	13%
Adapt_Peggy	29.75%
Adapt_Tanya	38.25%

Results and Conclusions

Results and scores

Sub task results	
Query	Percentage
Adapt_Archie	19.50%
Adapt_Jack	15.75%
Adapt_Max	12%
Adapt_Peggy	12%
Adapt_Tanya	27%

Results and Conclusions

Future plans

- **Text analysis** Separate handling of questions related to individual characters; perform a more detailed analysis of subscripts.
- **Voice detection** Improve accuracy of voice recognition, current results with SincNet tools are not accurate enough to be useful.
- **GAN** Provide more accurate augmented images by changing angles and lighting.

Questions